ЯЗЫКИ НАРОДОВ ЗАРУБЕЖНЫХ СТРАН (С УКАЗАНИЕМ КОНКРЕТНОГО ЯЗЫКА ИЛИ ГРУППЫ ЯЗЫКОВ) / LANGUAGES OF PEOPLES OF FOREIGN COUNTRIES (INDICATING A SPECIFIC LANGUAGE OR GROUP OF LANGUAGES)

DOI: https://doi.org/10.60797/RULB.2024.60.25

ON SOME ASPECTS OF CORPUS APPROACH TO IDIOMS

Research article

Ivanova E.V.[1, *]
[1] ORCID : 0000-0002-1990-3061;
[1] St. Petersburg State University, St. Petersburg, Russian Federation

* Corresponding author (e.v.ivanova[at]spbu.ru)

Abstract

The article is aimed at examining the lines of phraseological studies that can gain advantage from including corpus data into the focus of researchers' attention. The article relies on the corpus, semantic, structural and conceptual analysis of idioms containing the names of five body parts / organs as their components. The frequency of these idioms, their usage with possessive pronouns as variable components and the relevance of corpus data for conceptual modelling are considered. The basic results of the undertaken research concern the irrefutable value of corpus data for practical, i.e. lexicographic and teaching, purposes in the description of idioms, as well as for its theoretical importance for the simulation of conceptual constructs represented by idioms.

Keywords: idiom, corpus, corpus analysis, n-gram, frequency.

О НЕКОТОРЫХ АСПЕКТАХ КОРПУСНОГО ПОДХОДА К ИДИОМАМ

Научная статья

Иванова Е.В.[1, *]
[1] ORCID : 0000-0002-1990-3061;
[1] Санкт-Петербургский государственный университет, Санкт-Петербург, Российская Федерация

* Корреспондирующий автор (e.v.ivanova[at]spbu.ru)

Аннотация

Цель статьи заключается в рассмотрении направлений исследования фразеологических единиц, которые выигрывают от привлечения корпусных данных в поле зрения исследователей. Статья основывается на корпусном, семантическом, структурном и концептуальном анализе идиом, содержащих в качестве своих компонентов наименования пяти частей тела / органов человека. В статье производится анализ частотности данных идиом, использование притяжательных местоимений как переменных компонентов в их структуре и значимость корпусных данных для концептуального моделирования. Основные результаты проведенного исследования касаются бесспорной важности корпусных данных для практических, т.е. лексикографических и обучающих, целей при описании идиом, а также для теоретических целей, связанных с моделированием концептуальных конструктов, репрезентированных идиомами.

Ключевые слова: идиома, корпус, корпусный анализ, n-грамма, частотность.

Introduction

Nowadays, corpus analysis is widely used in many areas of linguistics, providing the researchers with the indispensable data about the actual usage of language signs in speech. Phraseology also experiences the advantages of using corpora as tools for idiom studies. Corpus analysis allows researchers to better define the actual meaning of idioms [1], to reveal their modifications in speech and to specify their typical concordances [2]. Corpus approach also makes it possible to identify the most frequent idioms for optimizing the material for English studies [5], as well as to improve the quality of translation [7], as a consequence helping to solve some practical and pragmatic issues. However, some difficulties in collecting corpus data about idioms are encountered, due to the relatively rare usage of idioms in texts, their frequent similarity in structure to non-idiomatic combinations, and their length, which may turn them into unidentifiable n-grams [9]. Nevertheless, the potential advantages for the research far outweigh the confronted obstacles that are in most cases successfully dealt with.

Research methods and principles

Corpus analysis is based on the Corpus of Contemporary American English (COCA) [10], considered to be the most sizeable and large-scale corpus, including more than one billion words, that function in eight genres of informal and formal texts. It is considered to be the most widely used and reliable corpus of the English language [5], [11]. Idioms are defined in this article as central units of phraseology, based on imagery and characterized by the irreducibility of their meaning to the summed up meanings of the constituent components. Idioms are collected from the Oxford Dictionary of Idioms [12]. This dictionary contains idioms used in the main English areal variants, so the basic principle underlying their collection was the selection of those that are used in American English and those that belong to all areal variants, including that of American English. Thus, idioms like *get your head down*, marked as British, were excluded. The Oxford Dictionary of Idioms does not contain a vast assembly of idioms, so it was assumed that its compilers had chosen the most widely used and important ones, relying on their own introspection and textual evidence.

The article focuses on the idioms with the components *head, eyes/eye, ears/ear, nose* and *heart*. Corpus analysis is employed along with the structural, semantic and conceptual analysis.

Main results

The conducted research produces the following results:

1. The corpus analysis even of a relatively small number of idioms clearly illustrates the necessity for lexicographers to take into account the evidence of idiom frequency in the corpus, which may result in the revision of the material included into concise dictionaries of idioms. The corpus data of idiom frequencies also provide valuable information for the compilers of text books and teachers of English. Both provisions have been made in the scientific literature before, as follows from the introduction to this article, but highlighting them once again might contribute to their practical realization.

2. The variable frequency of possessive pronouns as changeable components of a group of idioms supplies additional information about the functioning of such idioms in texts. This information may also be considered in the compilation of dictionaries and teaching materials, especially in those cases when tendencies of this or that pronoun prevalence are particularly distinct.

3. The corpus analysis is indispensable for the area of conceptual modelling based on the semantics of idioms. Modelling of various fragments of phraseological pictures of the world or of phraseological concepts has been conducted intensively in cognitive and cultural studies of phraseology for many years now (Cf. the works by A.A. Butina [2], A.V. Moskalenko [6], E. Sereda [8] and hundreds of others), but inclusion of corpus data in the process of modelling has not yet been realized to any significant extent. The different frequencies of idioms representing this or that concept supply the information about the basic components of meaning important for the native speakers.

Discussion

Names of body parts and organs constitute a very important segment of lexis in any language, reflecting the vast interest of people in themselves and their bodies. The essentiality of body code for culture and language is thoroughly and scrupulously argued by M.L. Kovshova [4]. But the significance of each body part / organ and hence the frequency of its name in speech varies, which is markedly illustrated by corpus data. These data for the body parts names, chosen for the analysis, are as follows:

eyes/eye – 349526 entries
head – 340401 entries
heart – 193144 entries
ears/ear – 60192 entries
nose – 41465 entries

As we can easily see, the nouns *eyes/eye* and *head* clearly prevail in their frequency over the other members of the selected group, which allows us to make a conclusion about the higher relevance of body parts they designate to the native speakers. This higher relevance is accounted by the fact, that eyes provide people with the predominant share of information about the world and the head accommodates the brain for processing this information and guiding people through their lives.

If we turn to the number of idioms containing the selected nouns, which the Oxford Dictionary of Idioms lists, we can also note the prevalence of the first two nouns as components of idioms, though the difference is not that striking:

eyes/eye – 30 idioms
head – 29 idioms
heart – 17 idioms
ears/ear – 17 idioms
nose – 15 idioms

The explanation for this prevalence is similar to the one given above in connection with the variable number of entries in the corpus.

The common feature of all the idioms under consideration is the predominant number of verbal units, which demonstrates the observation of body parts as involved in a situation with the further metaphorical perception of this situation. E.g., out of 29 idioms with *head*, 26 are verbal, out of 15 idioms with *nose* 11 are verbal, etc.

The selected idioms vary significantly in their frequency in the corpus. This variability can be traced both among the idioms with different names of body parts as their components and among the idioms having the same name of the body part.

The idiom *turn a blind eye* has 1052 entries, *wear your heart on your sleeve* - 49 entries, while *lead someone by the nose* – only 8. The adverbial idioms *in your heart of hearts* and *with your nose in the air* have 433 and 10 entries correspondingly. The idiom *have your heart in your mouth* cannot be traced in the corpus at all. As we can see, the difference can be quite dramatic and provides convincing data about the actual usage of the idioms incorporated into the dictionary.

The same concerns the idioms with the identical name of a body part. E.g., as shown above, the idiom *wear your heart on your sleeve* has 49 entries, while *in your heart of hearts* has 433. On the other hand, the idioms *with your nose in the air* and *lead someone by the nose* both demonstrate very low frequencies of 18 and 8 entries accordingly, though the former exceeds the latter in its frequency more than two times.

The search for idioms in the corpus can turn out to be more complicated, if idioms have homonyms in the form of independent combinations of words. Such idioms as *on the nose* (precisely), *hold (or put) a gun (or a pistol) to someone's head* (force someone to do something by using threats), *keep your head above water* (avoid succumbing to difficulties) have corresponding homonyms with the literal meanings.

…he punched and choked his girlfriend, then forced his young daughters to hold a gun to the woman's head.

Did the media hold a gun to Rick Perry's head to tell us how he would base foreign policy …

In the situation described by the first sentence, a gun was literally brought to a person's head, while in the second situation no gun as such was employed.

In such cases, it is inevitably required to sort out the accumulated corpus data manually, separating idioms from non-fixed word combinations. In this connection, it is possible to say that the idioms with the component *heart* differ from all the idioms with other components considered here, as none of them have homonyms with a literal meaning.

Another line of research is connected with the usage of possessive pronouns in the idioms that contain a variable component. The idiom *from the bottom of your heart* has 487 entries with the possessive pronoun *my*, 22 with *his*, 12 with *your*, 6 with *her*, 2 with *their* and 2 with the pronoun *our.* In the idiom *in your heart of hearts* the difference in the frequency is not so dramatic: *my* (131), *your* (101), *their* (76), *his* (68), *her* (36), *our* (21). Nevertheless, the undoubtable high frequency of *my* can most probably be explained by the semantics of the idiom and the apparent connection of heart for the native speaker with the soul and inner feelings, hence the fact that it is more natural for a person to talk about his own feelings, than those of others. The infrequently used idiom *with your head in the clouds,* on the other hand, displays a relatively insignificant difference in the frequency of alternating possessive pronouns: *your* – 9, *his* – 8, *my* – 5, *her* – 4, *their* – 1, *our* – 0. We can conclude that this idiom is used in connection with one individual, not a group of people, but otherwise the prevalence of *your* and *his* over *my* and *her* cannot be accounted for by any usage preference typical of native speakers.

If we want to resort to the conceptual modelling based on idioms and describe the perception of the world inherent to the native speakers and reflected in the images of the idioms with the names of body parts / organs, corpus data will also be very useful. E.g., in modelling the concept of heart represented by idioms, we conclude that the heart is perceived as a container with a bottom (*from the bottom of your heart*), that it can be made of different materials (*heart of gold; heart of oak: heart of stone*), and can be located in various places (*wear your heart on your sleeve; have (or put) your heart in; have your heart in your mouth; have your heart in the right place*). But apart from describing the imagery side of the concept and the involved metaphors with metonyms, it stands to reason to include the information about the prevailing possessive pronouns and the frequencies of idioms usage, extracted from the corpus. Thus, the prevalence of the pronoun *my* indicates the importance of the first-person singular attribution of the meaning of the idioms, indirectly implying the connection of the heart image to this particular possessor, which provides additional information to the model of the concept. Likewise, the high frequency of the idiom *from the bottom of your heart* (with sincere feeling) signifies the importance of the connection between “heart” and “feeling”, traced not only from the semantics of the idiom, but from its wide usage as well.

Conclusion

Corpus analysis opens up new prospects for phraseological studies and deserves the attention on the part of researchers in this respect. It provides valuable data about the functioning of idioms in speech and permits the researchers to classify idioms according to their frequencies. This classification has far-reaching results, concerning practical (lexicographic, teaching) and theoretical (conceptual-semantic modelling) areas. A certain discrepancy between the idioms selected for the dictionary and their actual frequency in speech, traced even for a very limited group, highlights the necessity for lexicographers and teachers to take corpus data into account. The same refers to the alternating frequency of possessive pronouns as variable components of idioms. This aspect of idiom functioning in speech has never been paid much attention to, but besides the pragmatic area it appears to be important for conceptual modelling in cognitive and cultural linguistic studies. These studies are only beginning to incorporate corpus data into their sphere of research, but the prospects that are opening here are vast and comprehensive.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Баранов А.Н. Аспекты теории фразеологии / А.Н. Баранов, Д.О. Добровольский. — М.: Знак, 2008. — 656 с.

2. Бутина А.А. Конwцепты ЦАТ и ДОГ в английской языковой картине мира: автореф. дис. … канд. филол. наук / Бутина Александра Анатольевна. — СПб, 2012. — 22 с.

3. Добровольский Д.О. Корпусы текстов и двуязычная фразеография / Д.О. Добровольский // Вестник Новосибирского государственного педагогического университета. — 2015. — № 5 (27). — С. 23–37.

4. Ковшова М.Л. Лингвокультурологическое направление во фразеологии: коды культуры / М.Л. Ковшова. — М.: ЛЕНАНД, 2016. — 453 с.

5. Комарова И.А. Исследование английской фразеологии с помощью подходов корпусной лингвистики / И.А. Комарова, М.С. Коган // Компьютерная лингвистика и вычислительные онтологии. — 2019. — Вып. 3. — С. 40–49.

6. Москаленко А.В. Концепт «Птица» в английской фразеологической картине мира: автореф. дис. … канд. филол. наук / Москаленко Анна Валерьевна. — СПб, 2015. — 21 с.

7. Пивоварова Е.В. Корпусный анализ как инструмент для выявления семантических изменений у фразеологизмов — «ложных друзей» переводчика / Е.В. Пивоварова // Филологические науки. Вопросы теории и практики. — 2020. — Т. 13. — Вып. 9. — С. 312–320.

8. Середа Е. Представление концепта «Death» в английской фразеологической картине мира: автореф. дис. … канд. филол. наук / Е. Середа. — СПб, 2016. — 22 с.

9. Cheng W. Exploring Corpus Linguistics: Language in Action / W. Cheng. — London: Routledge, 2012. — 256 p.

10. Corpus of Contemporary American English. — URL: https://www.english-corpora.org/coca/ (accessed: 04.11.2024).

11. Lindquist H. Corpus linguistics and the description of English / H. Lindquist. — Edinburgh: Edinburgh University Press, 2009. — 219 p.

12. Oxford Dictionary of Idioms. 2nd. Siefring J. (ed.). — Oxford: Oxford University Press, 2004. — 340 p.

Список литературы на английском языке / References in English

1. Baranov A.N. Aspekty teorii frazeologii [Aspects of phraseology theory] / A.N. Baranov, D.O. Dobrovol'skiy. — M.: Znak, 2008. — 656 p. [in Russian]

2. Butina A.A. Kontsepty CAT i DOG v angliyskoy yazykovoy kartine mira [Concepts of CAT and DOG in the English language picture of the world]: abstract dis. ... of PhD in Philology / Butina Alexandra Anatolyevna. — SPb, 2012. — 22 p. [in Russian]

3. Dobrovol'skiy D.O. Korpusy tekstov i dvuyazychnaya frazeografiya [Text corpora and two-language phraseology] / D.O. Dobrovol'skiy // Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta [Bulletin of Novosibirsk State Pedagogical University]. — 2015. — № 5 (27). — P. 23–37. [in Russian]

4. Kovshova M.L. Lingvokul'turologicheskoe napravlenie vo frazeologii: kody kul'tury [Linguistic cultural studies in phraseology: cultural codes] / M.L. Kovshova. — M.: LENAND, 2016. — 453 p. [in Russian]

5. Komarova I.A. Issledovanie angliyskoy frazeologii s pomoshch'yu podkhodov korpusnoy lingvistiki [Research of English phraseology with the corpus linguistics approach] / I.A. Komarova, M.S. Kogan // Komp'yuternaya lingvistika i vychislitel'nye ontologii [Computational Linguistics and Computer Science Ontologies]. — 2019. — Iss. 3. — P. 40–49. [in Russian]

6. Moskalenko A.V. Kontsept «Ptitsa» v angliyskoy frazeologicheskoy kartine mira [Concept “Bird” in the English phraseological picture of the world]: abstract dis. ... of PhD in Philology / Moskalenko Anna Valeryevna. — SPb, 2015. — 21 p. [in Russian]

7. Pivovarova E.V. Korpusnyy analiz kak instrument dlya vyyavleniya semanticheskikh izmeneniy u frazeologizmov — «lozhnykh druzey» perevodchika [Corpus analysis as a tool for revealing semantic changes in phraseological units — “false friends” of a translator] / E.V. Pivovarova // Filologicheskie nauki. Voprosy teorii i praktiki [Studies in Philology. Theoretical and Practical Issues] —2020. — Vol. 13. — Iss. 9. — P. 312–320. [in Russian]

8. Sereda E. Predstavlenie kontsepta “Death” v angliyskoy frazeologicheskoy kartine mira [Representation of the concept “Death” in the English phraseological picture of the world]: thesis of the PhD dissertation in philology] / E. Sereda — SPb, 2016. — 22 p. [in Russian]

9. Cheng W. Exploring Corpus Linguistics: Language in Action / W. Cheng. — London: Routledge, 2012. — 256 p.

10. Corpus of Contemporary American English. — URL: https://www.english-corpora.org/coca/ (accessed: 04.11.2024).

11. Lindquist H. Corpus linguistics and the description of English / H. Lindquist. — Edinburgh: Edinburgh University Press, 2009. — 219 p.

12. Oxford Dictionary of Idioms. 2nd. Siefring J. (ed.). — Oxford: Oxford University Press, 2004. — 340 p.